




Pronóstico de la tendencia de COVID-19 para 2021 en el noroeste de Argentina

COVID-19 prediction of tendency for 2021 in northwestern Argentina

Eduardo Agustín Mendoza^I , Octavio Bruzzone^{II} , María Julia Dantur Juri^{III,IV} 

RESUMEN: Usando un modelo de regresión polinomial con retraso, que empleó datos de COVID-19 de 2020 con ausencia de vacunas, se realizó la predicción de COVID-19 en un escenario con administración de vacunas para Tucumán en 2021. La modelación incluyó la identificación de un punto de quiebre de contagios entre ambas series con la mejor correlación. Previamente, se indicó por medio de correlación cruzada el lag que sirvió para obtener el menor error entre los valores esperados y los observados. La validación del modelo fue realizada con datos reales. En 21 días fueron predichos 18.640 casos de COVID-19 de 20.400 casos informados. El pico máximo de COVID-19 fue estimado 21 días antes con la intensidad esperada.

Palabras clave: Predicción. Modelo. COVID-19. Vacunas.

^IFundación Miguel Lillo, Instituto de Ecología, Comportamiento y Conservación, Biología Integrativa – San Miguel de Tucumán (Tucumán), Argentina.

^{II}Unidad Ejecutora Instituto de Investigaciones Forestales y Agropecuarias Bariloche, (Instituto Nacional de Tecnología Agropecuaria/ Consejo Nacional de Investigaciones Científicas y Técnicas) – San Carlos de Bariloche, Argentina.

^{III}Unidad Ejecutora Lillo (Consejo Nacional de Investigaciones Científicas y Técnicas-Fundación Miguel Lillo) – San Miguel de Tucumán (Tucumán), Argentina.

^{IV}Unidad de Microscopía Espectral (Universidad Nacional de Tucumán, Sistema Provincial de Salud, Ministerio de Salud de Tucumán) – San Miguel de Tucumán, (Tucumán), Argentina.

Autor correspondiente: Eduardo Agustín Mendoza, Instituto de Ecología, Comportamiento y Conservación, Biología Integrativa, Miguel Lillo 251, (CP 4000), San Miguel de Tucumán (Tucumán), Argentina. E-mail: eamendoza@lillo.org.ar.

Conflicto de intereses: nada a declarar – **Fuente de financiamiento:** ninguna.

ABSTRACT: Using a lagged polynomial regression model, which used COVID-19 data from 2020 with no vaccines, the prediction of COVID-19 was performed in a scenario with vaccine administration for Tucumán in 2021. The modeling included the identification of a contagion breaking point between both series with the best correlation. Previously, the lag that served to obtain the smallest error between the expected and observed values was indicated by means of cross correlation. The validation of the model was carried out with real data. In 21 days, 18,640 COVID-19 cases out of 20,400 reported cases were predicted. The maximum peak of COVID-19 was estimated 21 days in advance with the expected intensity.

Keywords: Forecasting. Model. COVID-19. Vaccines.

INTRODUCCIÓN

En Marzo de 2020, la Organización Mundial de la Salud (OMS) declaró a la enfermedad coronavirus (COVID-19) como pandemia¹. Instó a activar diversos protocolos para contener su propagación². En Argentina, el primer caso fue detectado en marzo de 2020 en Buenos Aires, declarándose cuarentena obligatoria por Decreto de Necesidad y Urgencia³.

A principios de 2021, no se habían administrado vacunas a la población y luego de la reapertura de actividades, comenzó la segunda ola de COVID-19.

El objetivo fue predecir la tendencia de los casos COVID-19 durante 2021 para un escenario con administración de vacunas y su pico máximo, estudiando el comportamiento estadístico de los datos de COVID-19 en 2020 sin la aplicación de vacunas.

MÉTODOS

El estudio se llevó a cabo en la provincia de Tucumán, en el noroeste de Argentina, resultando elegida debido a la falta de predicción de casos de COVID-19 y por ser la segunda provincia más densamente poblada del país, con un reporte de 1.338.523 habitantes⁴.

La elaboración del modelo de predicción para casos de COVID-19 consistió en identificar en datos de COVID-19 de 2020 un lag de días t que mejor se correlacione con un lag de días t de COVID-19 de 2021, empleando como referencia un punto de quiebre de contagios en la primera serie. Con esa identidad lograda, se realizó una correlación cruzada entre esos lags, con la finalidad de encontrar el mejor retraso para ajustar los datos con un modelo polinomial de regresión con retraso y predecir a la tendencia de COVID-19 actual.

Transformación de los datos: para estabilizar el promedio y reducir la tendencia se usó diferenciación de orden uno. El valor de p se realizó con el estadístico t con $n-2$ grados de libertad y con n basado en el número de muestras que se superponen en las correlaciones cruzadas. El análisis fue realizado con Past 3.22^{5,6}.

Se emplearon dos conjuntos de datos de COVID-19, los que fueron publicados diariamente por el Ministerio de Salud Pública de la provincia de Tucumán (MSPT)⁷. El primer conjunto desde el 18/03/2020 hasta 27/11/2020. El segundo conjunto desde el 19/03/2021 hasta 20/05/2021. Se elaboró una matriz de lags para COVID-19 de 2020 con diferentes cantidades de días de longitud. Previamente, se obtuvieron las fechas de inicio y final de los lags. Se obtuvieron con base en un punto de quiebre de contagios, indicado por un incremento de 50% de la totalidad de casos reportados antes del pico de COVID-19 en 2020. Se empleó una media móvil de 15 días. La longitud en t días de los lags (l) fueron exploradas a 30, 35, 40 y 45 días.

La identificación del lag de COVID-19 2020 para elaborar el set de entrenamiento se determinó por correlación de Pearson (r_p) con $p > 0,05$ con el lag de COVID-19 de 2021. Identificado el lag, se realizó correlación cruzada (r_d) con $p > 0,05$ entre ellos. De esta manera se obtuvo la ubicación en la serie predictiva y_i (COVID-19 2020) para su mejor retraso m . En el Apéndice 1, se indica mediante un diagrama de flujo a la metodología empleada en detalles.

Modelo empleado: el lag de COVID-19 de 2020 identificado en el retraso m junto al lag de datos de COVID-19 de 2021 fueron ajustados con un modelo de regresión polinomial con retraso. Se empleó este tipo de modelo debido a que los casos de COVID-19 son aleatorios y no lineales. El modelo de regresión Polinomial⁸ utilizado fue:

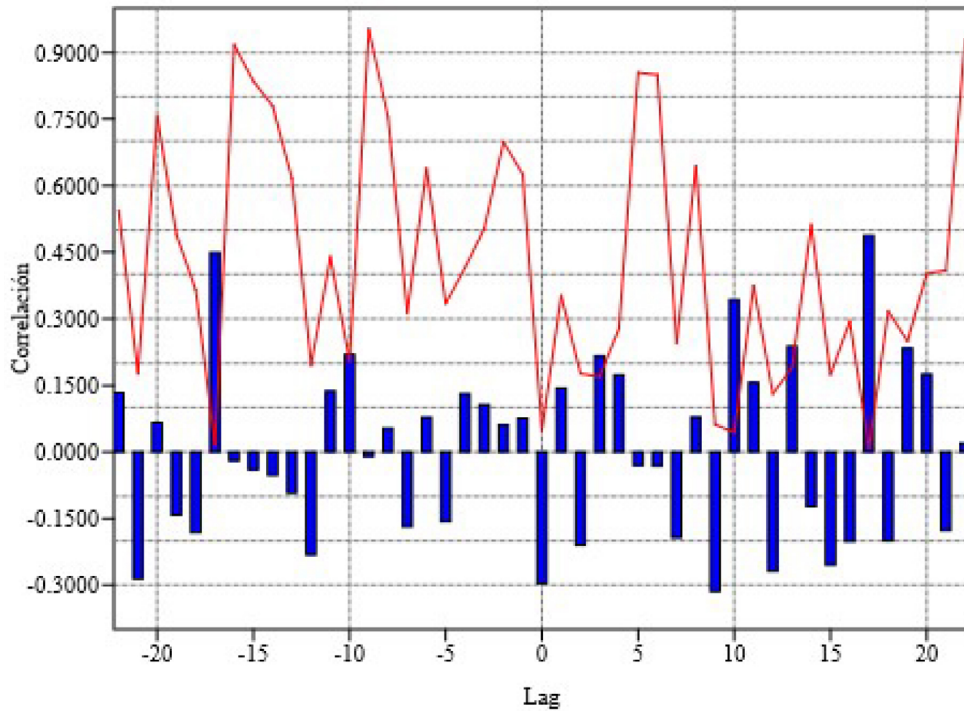
$$x_i = a + b y_{i-m} + c y_{i-m}^2 + \dots + e$$

En donde x_i representa los casos de COVID-19 diferenciados y predichos para 2021 en el día i , a , b , c son coeficientes del modelo polinomial, y es la serie predictiva de COVID-19 de 2020 del día i que mejor pronostica x en función de y para su mejor retraso $i-m$, mientras e representa el error estimado. El proceso fue invertible para la diferenciación realizada. La autocorrelación de los residuos del mejor modelo fue nula. La evaluación del modelo se realizó con datos reales de COVID-19, empleando el error de pronóstico porcentual promedio (MAPE). Un horizonte de pronóstico fue valorado posteriormente de idéntica forma.

RESULTADOS

Los resultados indicaron que el punto de quiebre de contagios de COVID-19 en 2020 fue el 25/09/2020 (692 casos), mientras que en los datos de 2021 fue el 13/05/2021 (723 casos).

El lag de datos COVID-19 en 2021 empleado para construir el modelo fue desde el 30/03/2021 hasta el 13/05/2021 y para el lag de COVID-19 en 2020 fue desde el 18/08/2020 al 01/10/2020. Entre las series se obtuvo $r_p = -0,296$, con $p = 0,04$, mientras que para un retraso de $m = 17$ días fue $r_d = 0,488$ con $p = 0,008$ (Figura 1). El modelo obtenido para ese retraso fue: $x_i = -1,935E-06 y_{i-m}^3 + 0,0006216 y_{i-m}^2 + 0,02296 y_{i-m} + 11,44$, con $R^2 = 0,315$, $F = 36,8$, $p = 0,026$ (Figura 2). La autocorrelación de los residuales fue nula.



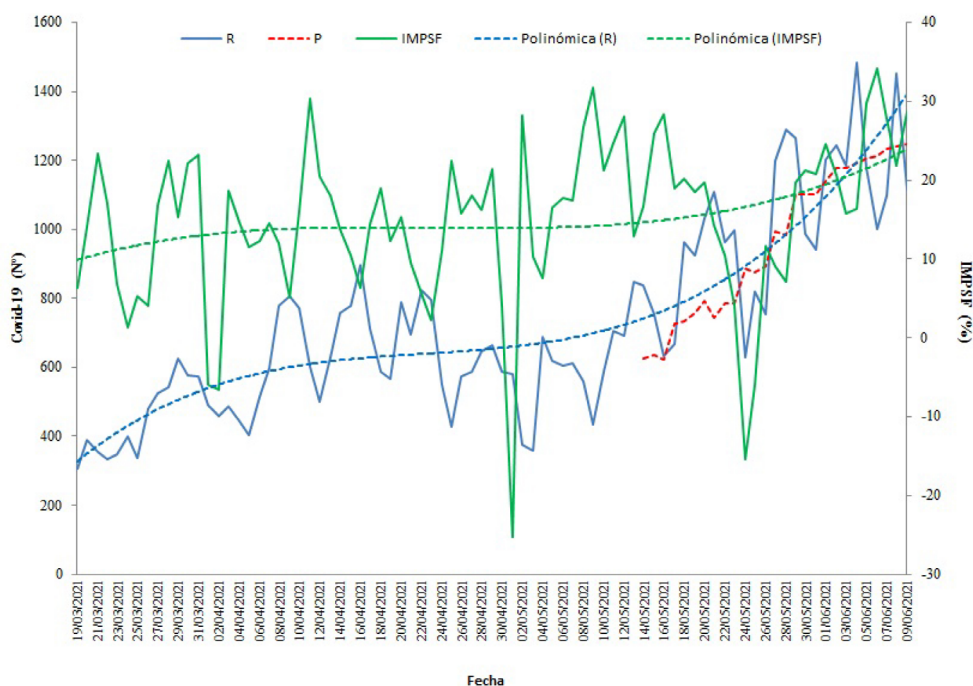
La línea roja indica probabilidad. El lag 10 fue significativo con residuales de modelo no nulo.

Figura 1. Correlograma entre el lag de entrenamiento de COVID-19 2020 y el lag de COVID-19 de 2021.

Los 17 días predichos fueron desde el 14/05/2021 hasta el 30/05/2021 ($n=17$) (Figura 2) con 14.042 casos predichos de 15.824 reportados, su MAPE fue de 11,3. En la máxima variabilidad de los casos de COVID-19 desde el 31/05/2021 hasta el 03/06/21 se acumularon 4.576 casos predichos sobre 4.598 casos reportados con MAPE de 0,47. El pico máximo de COVID-19 a 21 días fue estimado para el 03/06/2021 con 1.200 casos y ocurrió el 04/06/202 con 1.485 casos.

DISCUSIÓN

Los resultados mostraron que el modelo subestimó la cantidad de sucesos ocurridos antes del 27/05/2021, momento en que se instalaron restricciones sociales estrictas⁹ y se acompañaron a los casos reales hacia el pico máximo cuando se instalaron las restricciones. Es posible que la subestimación se encuentre influenciada por relajación de la sociedad en cuanto a la administración de vacunas. Antes del inicio del modelado el 22/04/2021, la campaña de vacunación acumuló 230.000 dosis aplicadas¹⁰. Mientras dos días después del primer



Las líneas azul R y Polinómica (R) corresponden a datos reales de COVID-19. La línea roja P corresponde a datos de COVID-19 predichos con MRPR. La línea verde es un Índice de Movimiento de Personas en Supermercados y Farmacias (IMPSF) en porcentaje para Tucumán según su punto de referencia igual acero. Todas las líneas quebradas corresponden a polinomios de grado tres.

Figura 2. Datos reales y pronosticados de casos de COVID-19 en 2021 en Tucumán.

pico de COVID-19, el 06/05/2021, se acumuló la aplicación de 306.000 dosis¹¹. Otra forma de subestimación del modelo sería la ausencia de restricciones sociales. Comparamos de manera conjunta el incremento de los casos de COVID-19 informados por el MSPT, los predichos y un Índice de Movimiento de Personas en Supermercados y Farmacias¹² y observamos que se comportaron de manera similar (Figura 2).

La precisión del modelo es similar al de otras investigaciones informadas, como el calculado con modelo de supervivencia y convolución, parsimonioso y robusto¹³. La duración de la predicción obtenida es similar a la lograda con el modelo susceptible-expuesto-infectioso-recuperado extendido¹⁴.

El modelo que presentamos pudo predecir la tendencia en la dinámica de los casos esperados de COVID-19 hacia el pico máximo. Sin embargo, solamente pudo predecir el pico de COVID-19 para el día 03 de junio, ocurriendo realmente dos picos de COVID-19 en 2021, uno en el día 04/06/21 y el otro en el 08/06/21.

En conclusión, destacamos que las tendencias de los casos COVID-19 en 2021 en Tucumán pudieron predecirse analizando el comportamiento estadístico de la primera ola de COVID-19 ocurrida en 2020.

REFERENCIAS

1. World Health Organization. Coronavirus disease 2019 (COVID-19) situation report, 51. Ginebra: World Health Organization; 2020. [accedido el 5 jun. 2021]. Disponible el: <https://apps.who.int/iris/bitstream/handle/10665/331475/nCoVsitrep11Mar2020-eng.pdf?sequence=1&isAllowed=y>
2. World Health Organization. Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). Ginebra: World Health Organization; 2020. [accedido el 5 jun. 2021]. Disponible el: <https://www.who.int/docs/default-source/coronavirus/who-china-joint-mission-on-covid-19-final-report.pdf>
3. Argentina. Ministerio de Salud. Decreto de Necesidad y Urgencia 260/2020 [Internet]. 2020. [accedido el 5 jun. 2021]. Disponible el: <https://www.argentina.gob.ar/coronavirus/dnu>
4. Instituto Nacional de Estadística y Censo. República Argentina. Censo Nacional de Población, Hogares y Viviendas 2010. [Internet]. 2020 [accedido el 21 may. 2021]. Disponible el: <https://www.indec.gob.ar/indec/web/Nivel4-CensoProvincia-3-999-90-000-2010>
5. Hammer Ø, Harper DAT, Ryan PD. 2001. PAST: Paleontological software package for education and data analysis. *Paleontological Electronica* 4(1):9. Disponible el: <https://www.nhm.uio.no/english/research/infrastructure/past/>
6. Covid19-prediction. COVID-19 prediction of tendency for 2021 in northwestern Argentina. Disponible el: <https://github.com/Agustino216/Covid19-prediction>
7. Gobierno de Tucumán. Ministerio de Salud Pública. [Internet]. 2020. [accedido el 29 may 2021]. Disponible el: <https://msptucuman.gov.ar/category/noticias/>
8. Cromwell JB, Labys WA, Hannan MJ, Terraza M. Multivariate tests for time series models. USA: SAGE University paper.
9. Gobierno de Tucumán. Comité Operativo de Emergencia de Tucumán. 2021. [Internet] [accedido el 22 may 2021]. Disponible el: https://coe.tucuman.gov.ar/recursos/documentos/archivos/archivo_333_20210522110758.pdf
10. Gobierno de Tucumán. Ministerio de Salud Pública. Llegaron 24600 dosis de Sputnik V a la provincia [Internet]. 2021. [accedido el 31 oct. 2021]. Disponible el: <https://vacunartuc.gob.ar/llegaron-24600-dosis-de-sputnik-v-a-la-provincia/>
11. Gobierno de Tucumán. Ministerio de Salud Pública. Desde el inicio de la campaña Tucumán lleva aplicadas 306.833 dosis de vacunas contra el Covid-19. [Internet]. 2021. [accedido el 31 oct. 2021]. Disponible el: <https://vacunartuc.gob.ar/desde-el-inicio-de-la-campana-tucuman-lleva-aplicadas-306-833-dosis-de-vacunas-contra-el-covid-19/>
12. Google COVID-19 Community Mobility Reports. [Internet]. 2021. [accedido el 01 nov. 2021]. Disponible el: <https://www.google.com/covid19/mobility/>
13. Wang Q, Xie S, Wang Y, Zeng D. Survival-convolution models for predicting COVID-19 cases and assessing effects of mitigation strategies. *Front Public Health* 2020; 8: 325. <https://doi.org/10.1101/2020.04.16.20067306>
14. Ghostine R, Gharamti M, Hassrouny S, Hoteit I. An extended SEIR model with vaccination for forecasting the COVID-19 pandemic in Saudi Arabia using an ensemble kalman filter. *Mathematics* 2021; 9(6): 636. <https://doi.org/10.3390/math9060636>

Recibido el: 02/09/2021

Revisado el: 05/11/2021

Aceptado el: 07/12/2021

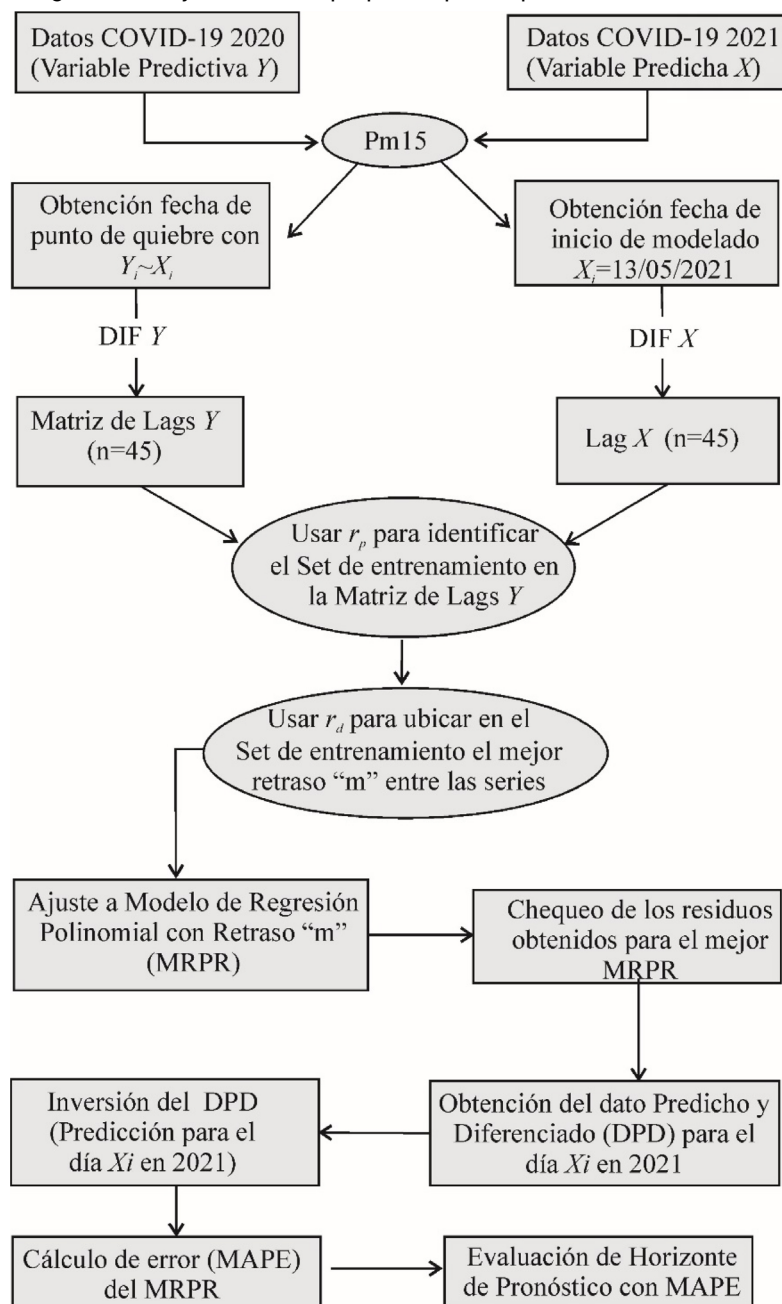
Preprint: 13/12/2021

<https://preprints.scielo.org/index.php/scielo/preprint/view/3346>

Comité de ética: Se trabajó con reportes diarios de casos de COVID-19 publicados por el Ministerio de Salud de Tucumán, siendo los mismos de libre acceso, sin necesidad de la aprobación de un Comité de Ética en Investigación.

Contribuciones de los autores: Mendoza, E.A.: Administración del proyecto, Análisis formal, Conceptuación, Consultoría de datos, Escrita – primera redacción, Escrita – revisión y edición, Investigación, Metodología, Software, Supervisión, Validación. Bruzzone, O.: Administración del proyecto, Escrita – revisión y edición, Supervisión, Validación, Visualización. Juri, M.J.D.: Administración del proyecto, Escrita – revisión y edición, Supervisión, Validación, Visualización.

Apéndice 1. Diagrama de flujo del modelo propuesto para la predicción de COVID-19 en Tucumán.



MAPE: error de pronóstico porcentual promedio, Pm: promedio móvil de 15 días, DIF: Diferenciación de datos de orden uno.